Claims

What is claimed is:

1.   An information retrieval system, comprising:

a hierarchal analysis component that receives a query and processes probabilities associated with N categories, each category having one or more topics, N being an integer; and

an interactive component that provides feedback derived from the query and the probabilities associated with the N categories and the one or more topics, the feedback being utilized to determine at least one category of the N categories to facilitate retrieval of at least one of the one or more topics.

2.   The system of claim 1, further comprising an automatic classifier construction component that builds a top-level classifier for the N categories and a sublevel classifier for each category of the one or more topics associated with the N categories.

3.   The system of claim 2, wherein the classifiers are provided by at least one of a Support Vector Machine, Naive Bayes, Bayes Net, decision tree, similarity-based, vector-based and a Bayesian-based classification model.

4.   The system of claim 3, wherein the automatic classifier construction component employs a learning model to build the classifiers.

5.   The system of claim 4, wherein the learning model is associated with a Support Vector Machine and employs Sequential Minimal Optimization (SMO) to train the classifiers.

6. The system of claim 4, further comprising a data structure that includes a mapping of I possible queries and one or more associated topics, I being an integer, to enable learning for the classifiers.

7. The system of claim 6, wherein the data structure is updated *via* at least one of implicit and explicit user actions associated with a query to facilitate improved learning models.

8. The system of claim 6, wherein the data structure is centrally located to enable monitoring of implicit and explicit user actions associated with queries from a plurality of users to facilitate improved learning models.

9. The system of claim 2, wherein the first classifier is employed to drive the sublevel classifiers at run time to form a hierarchical classification structure.

10. The system of claim 9, wherein the query and the first classifier are employed to determine the most likely of the N categories.

11. The system of claim 10, further comprising a context disambiguation component that utilizes the query and the first classifier to determine the feedback.

12. The system of claim 11, wherein the context disambiguation component utilizes the query and the feedback to drive the sublevel classifiers in order to determine a desired topic.

13. The system of claim 11, wherein the context disambiguation component further comprises a presentation component for interfacing to a user and an analytical component to facilitate feedback and decision-making related to the feedback.

14. The system of claim 13, wherein the analytical component includes a cost-benefit analysis considering the cost of the dialog with the information value of the dialog.

15. The system of claim 13, wherein the analytical component includes a decision analysis for determining the nature and quantity of a clarification dialog.

16. The system of claim 13, wherein the analytical component includes a computation of the value of information associated with feedback gained during a clarification dialog to guide the nature and quantity of the clarification dialog.

17. The system of claim 13, wherein the analytical component employs at least one of a rule-based policy and an expected utility policy that controls if and how dialog is invoked based on the distribution of probabilities assigned to topics at one or more layers of a classification scheme.

18. The system of claim 17, wherein the analytical component analyzes probabilistic weights associated with each category and related subtopic for determining feedback and presentation to the user.

19. The system of claim 17, wherein the analytical component analyzes probabilistic weights as a spread across each category and related subtopic for determining feedback and presentation to the user.

20. The system of claim 13, wherein the presentation component includes a ranked display of most likely N categories.

21. The system of claim 20, wherein at least one of the most likely N categories is selected to provide a ranked display of one or more topics.

22.   The system of claim 1, wherein information is retrieved as part of a help system.

23.   The system of claim 1, wherein information is retrieved from a network-based system.

24.   The system of claim 1, wherein the probabilities are determined *via* a hand-crafted analysis.

25.   The system of claim 1, further comprising L levels of N categories, each category having one or more topics, wherein L and N are integers.

26.   A computer-readable medium storing the computer-executable components of claim 1.

27.   A method providing information retrieval, comprising:
    determining probabilities associated with one or more categories associated with one or more associated topics;
    providing feedback that is derived from a query and the probabilities associated with the one or more categories and the one or more associated topics; and
    resolving at least one category of the one or more categories based upon the feedback to facilitate retrieval of at least one of the one or more associated topics.

28.   The method of claim 27, further comprising,
    building a top level classifier for the N categories and a sublevel classifier for each category of the one or more topics associated with the N categories.

29.   The method of claim 28, wherein the classifiers are at least one of a vector-based and a Bayesian-based model.

24

30. The method of claim 29, further comprising,

mapping I possible queries and associated topics within a data structure, I being an integer, to enable learning of the classifiers.

31. The method of claim 30, further comprising,

monitoring implicit and explicit user actions associated with a query to facilitate improved learning models.

32. The method of claim 30, further comprising,

monitoring a central data location for implicit and explicit user actions associated with queries from a plurality of users to facilitate improved learning models.

33. The method of claim 28, wherein the top-level classifier is employed to drive the sublevel classifiers at run time to form a hierarchical classification structure.

34. The method of claim 33, wherein the query and the top-level classifier are employed to determine the most likely of the N categories.

35. The method of claim 34, further comprising,

utilizing the query and the top-level classifier to determine the feedback.

36. The method of claim 35, further comprising,

utilizing the query and the feedback to drive the sublevel classifiers in order to determine a desired topic.

37.   The method of claim 27, further comprising,

utilizing at least one of a cost benefit analysis and a decision analysis for determining the feedback.

38.   The method of claim 35, further comprising,

utilizing rule-based policies and expected-utility policies for establishing probabilistic thresholds associated with the feedback.

39.   A system providing information retrieval, comprising:

means for determining probabilities associated with N categories, each category having one or more topics, N being an integer;

means for providing feedback that is derived from a query and the probabilities associated with the N categories and the one or more topics; and

means for determining at least one category of the N categories based upon the feedback to facilitate retrieval of at least one of the one or more topics.

40.   A computer-readable medium having stored thereon a data structure, comprising:

a plurality of query fields representing data associated with previous user interactions with a data base;

a plurality of topic fields representing possible topics associated with the query fields; and

at least one category field derived from the plurality of query and topic fields to enable information retrieval from the topics.

41.   A signal adapted to be transmitted between at least two processes, comprising:

an analysis component that receives a query *via* the signal and processes probabilities associated with N categories, each category having one or more topics, N being an integer; and

an interactive component that provides feedback *via* the signal derived from the query and the probabilities associated with the N categories and the one or more topics, the feedback being utilized to determine at least one category of the N categories to facilitate retrieval of at least one of the one or more topics.